Received: 22 March 2021 | Revised: 29 September 2021 | Accepted: 30 September 2021

DOI: 10.1002/bit.27950

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Kinetic and hybrid modeling for yeast astaxanthin production under uncertainty

Fernando Vega-Ramon[1] | Xianfeng Zhu[2] | Thomas R. Savage[1] | Panagiotis Petsagkourakis[3] | Keju Jing[2] | Dongda Zhang[1]

[1]Department of Chemical Engineering and Analytical Science, The University of Manchester, Manchester, UK

[2]Department of Chemical and Biochemical Engineering, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, China

[3]Department of Chemical Engineering, University College London, London, UK

Correspondence

Keju Jing, Department of Chemical and Biochemical Engineering, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China.
Email: jkj@xmu.edu.cn

Dongda Zhang, Department of Chemical Engineering and Analytical Science, The University of Manchester, The Mill, Sackville St. Manchester M1 3AL, UK.
Email: dongda.zhang@manchester.ac.uk

Abstract

Astaxanthin is a high-value compound commercially synthesized through *Xanthophyllomyces dendrorhous* fermentation. Using mixed sugars decomposed from biowastes for yeast fermentation provides a promising option to improve process sustainability. However, little effort has been made to investigate the effects of multiple sugars on *X. dendrorhous* biomass growth and astaxanthin production. Furthermore, the construction of a high-fidelity model is challenging due to the system's variability, also known as batch-to-batch variation. Two innovations are proposed in this study to address these challenges. First, a kinetic model was developed to compare process kinetics between the single sugar (glucose) based and the mixed sugar (glucose and sucrose) based fermentation methods. Then, the kinetic model parameters were modeled themselves as Gaussian processes, a probabilistic machine learning technique, to improve the accuracy and robustness of model predictions. We conclude that although the presence of sucrose does not affect the biomass growth kinetics, it introduces a competitive inhibitory mechanism that enhances astaxanthin accumulation by inducing adverse environmental conditions such as osmotic gradients. Moreover, the hybrid model was able to greatly reduce model simulation error and was particularly robust to uncertainty propagation. This study suggests the advantage of mixed sugar-based fermentation and provides a novel approach for bioprocess dynamic modeling.

KEYWORDS

batch operation, fermentation, hybrid modeling, mixed sugar, uncertainty analysis

1 | INTRODUCTION

Astaxanthin is a commercial carotenoid widely used as a natural red colorant in the food and cosmetic industries (Rodríguez-Sáiz et al., 2010). It is also an efficient antioxidant used in the nutraceutical and medical industries (Wan et al., 2014). The price of pure astaxanthin can be as high as $2,500 per kilogram (Lorenz & Cysewski, 2000). Although astaxanthin can be biologically synthesized by microagal *Haematococcus pluvialis* and yeast *Xanthophyllomyces dendrorhous*, the microalgal photo-production route is still at its infant stage due to a number of engineering challenges related to the low growth rate of algal cells and difficulty in designing large scale

photobioreactors (Zhang et al., 2016). As a result, the yeast fermentation route is more promising and has been applied for the commercial production of astaxanthin.

Within the last decades, extensive research has been carried out to investigate *X. dendrorhous* astaxanthin production. For example, a number of mutants have been genetically modified and screened to enhance biomass growth and astaxanthin production (Chi et al., 2015; Ukibe et al., 2008), different models have been developed and optimal substrate feeding strategies for astaxanthin synthesis have been proposed (Liu & Wu, 2007, 2008), the influence of oxygen and glucose on *X. dendrorhous* primary metabolism and astaxanthin accumulation has been explored (Liu & Wu, 2007), and upscaling of experiments has been executed at scales ranging from 2 L lab-scale bioreactors to 10,000 L fermenters (Rodríguez-Sáiz et al., 2010).

To further reduce process cost and improve sustainability, utilizing biowaste, for example, lignocellulosic biomass-derived sugars for fermentation is particularly attractive. However, the majority of these biowastes produce mixed sugars rather than a single compound after hydrolysis or other decomposition methods (Hallenbeck, 2012). As a result, it is important to investigate the effect of different types of sugars on *X. dendrorhous* biomass growth and astaxanthin production. A previous study has confirmed that amongst the three most commonly used sugars (sucrose, glucose, and fructose), fructose was the least favored option for carbon assimilation (An et al., 2001). However, little effort has been placed in the exploration of the relationships between different sugars on biomass growth and astaxanthin accumulation. As a result, the current study aims to investigate the process kinetics between single sugar (glucose) based fermentation and mixed sugar (glucose and sucrose) based fermentation. Sucrose was selected as the second sugar, rather than sugar monomers found in lignocellulosic hydrolysates, for the following reasons: (i) sucrose (disaccharide) was found to promote astaxanthin biosynthesis of *X. dendrorhous* in our previous experiments, whilst monosaccharides such as xylose, mannose, arabinose, or galactose did not increase astaxanthin accumulation when mixed with glucose, and (ii) the price of these monosaccharides is significantly higher than that of sucrose, so they have no cost advantage for industrial production.

An efficient approach for bioprocess kinetics investigation is to construct mathematical models. Bioprocess modeling in literature can be generally classified into two frameworks, namely structured and unstructured approaches. Unstructured approaches (such as the Monod-type kinetic models) provide a simplified representation of the bioprocess, considering the cell culture as a homogeneous biomass and describing the evolution of the process in terms of macroscopic state variables like substrate, product, and biomass concentration. These models are predominantly used for bioprocess systems engineering such as process monitoring, optimization, control, and reactor upscaling. On the other hand, structured approaches (such as cybernetic models) incorporate information about the microbial structure and physiology to obtain a mathematical description of the metabolism of the microorganism (Birol et al., 2002). These models can provide further insight into metabolic network activity and cellular regulation mechanisms, thus are often used for metabolic engineering and regulatory process simulation studies (Ramkrishna & Song, 2012). As the current research focuses on macroscopic bioprocess modeling, the unstructured modeling method is selected.

A particular challenge for bioprocess modeling is to account for bioprocess batch-to-batch variation. Bioprocesses often exhibit high variability due to their complex underlying process mechanisms. With the involvement of multiple phases (gas, liquid, and solid), the underlying process can behave in an unpredictable way over a broad range of time and length scales (Zhang et al., 2020). As a result, developing an accurate process model with high reliability (i.e., low uncertainty) is particularly challenging. Therefore, the second objective of this study is to explore a novel modeling approach, namely hybrid modeling, for bioprocess dynamic simulation under uncertainty. The yeast-based astaxanthin production process will be chosen as a case study to compare the performance of the hybrid model with the kinetic model.

Hybrid models are a new type of unstructured models that have been applied to several recent studies for biochemical process modeling (Cabaneros Lopez et al., 2021; Willis & von Stosch, 2017) and monitoring (Destro et al., 2020; Geinitz et al., 2020). These models incorporate a data-driven model into a conventional unstructured kinetic model to enhance the model's accuracy and predictive ability (Carinhas et al., 2011; von Stosch et al., 2014). In spite of recent success, most of the previous hybrid models were constructed using an artificial neural network (ANN) based data-driven model, which is not efficient for uncertainty estimation (their uncertainty is approximated using statistical methods such as bootstrapping) (Pinto et al., 2019). As a result, the current study will investigate a new avenue for hybrid model construction and uncertainty estimation.

2 | METHODOLOGY

2.1 | Introduction to experimental setup

The yeast *X. dendrorhous* strain was used in this study, which is an astaxanthin high-producing strain derived from ATCC 24230 by beta-ionone screening. The yeast strain was stored in frozen tubes at −80°C. The liquid medium for the single sugar (glucose) fermentation experiment was composed of 12 g glucose, 2 g $(NH_4)_2SO_4$, 1.5 g KH_2PO_4, 1.5 g $MgSO_4{\cdot}7H_2O$, 1 g NaCl, 2.5 g yeast extract (per liter), at pH 6.0. The mixed sugar medium added 6 g/L sucrose with other components the same as a single medium. A further experiment with the same culture medium but only using 6 g/L sucrose as the carbon source was also carried out. It is expected that by comparing the total biomass concentration and astaxanthin production in the mixed sugar experiment with those from the two single sugar experiments, one can deduce the interactions between sucrose and glucose. To avoid confusion, in this study, we refer to the glucose-based fermentation process as the "single sugar" process, as it is the most commonly used carbon source for industrial fermentation.

The inoculum for the shake-flask culture in the following experiments (5% v/v for all) was prepared by growing the cells in 250 mL flasks for 2 days, on a shaking table (250 rpm, 22°C). A folded 8-layer gauze was used as a filter device to ensure the oxygen demand of yeast and to prevent the risk of bacterial contamination. All fermentation experiments in 2000 ml Erlenmeyer flasks (400 ml filling volume) were run for 7 days at 22°C with a shaking speed of 250 rpm. Each experiment was repeated three times. The biomass, total sugar, and astaxanthin concentration were measured every 12 h. Biomass was measured by dry cell weight (DCW, g/L). Cells were collected in pre-weighed tubes by centrifugation at 8000 *g*, 8°C for 10 min. Cell pellets were washed twice with deionized water and dried at 100°C to constant weight. The supernatant was analyzed using a BioProfile 300 analyzer (Nova, USA) to measure the residual glucose concentration. Astaxanthin content was measured on an Agilent 1200 series HPLC system equipped with a UV detector (Agilent Technologies) and an Agilent reversed-phase HC-C18 column (4.6 × 250 mm, 5 µm). Pure methanol was used as the mobile phase at a flow rate of 1 ml/min and a column temperature of 30°C with a wavelength of 478 nm. It is worth noticing that only total sugar concentration can be measured in the mixed sugar experiment, thus it is not possible to directly measure the consumption of individual sugars.

2.2 | Kinetic model construction

2.2.1 | Single sugar kinetic model construction

Different unstructured kinetic models such as the Monod model, the Contois model, the Logistic model, and the hybrid Logistic-Monod model have been tested to simulate biomass growth (Y.-S. Liu & Wu, 2008; Xu, 2020; Zhang et al., 2015). The Contois model presented in Equation (1a) was found to give the best fitting result with the assumption that other process conditions, for example, oxygen concentration and temperature are maintained constant (Vatcheva et al., 2006).

$$\frac{dX}{dt} = \left(\mu_m \cdot \frac{S}{S + K_c \cdot X}\right) \cdot X - \mu_d \cdot X, \tag{1a}$$

$$\frac{dS}{dt} = -Y_S \cdot \left(\mu_m \cdot \frac{S}{S + K_c \cdot X}\right) \cdot X, \tag{1b}$$

$$\frac{dP}{dt} = \alpha \cdot \left(\mu_m \cdot \frac{S}{S + K_c \cdot X}\right) \cdot X + \beta \cdot X - k_d \cdot X^2, \tag{1c}$$

where μ_m is maximum specific growth rate, K_c is the substrate saturation constant, μ_d is specific decay rate, and X, S, and P are concentrations of biomass, substrate, and product, respectively. Y_S is the substrate yield coefficient. α is the astaxanthin growth-dependent yield coefficient, β is the growth-independent yield coefficient, and k_d is the specific consumption rate for astaxanthin.

When the carbon source is in excess ($S \gg K_C X$), the term $(S/S + K_c X) \approx 1$ and the yeast growth rate is independent of the substrate concentration. The Contois saturation constant K_c bears a similar physical significance to the saturation constant in the Monod equation; a small K_c indicates that the affinity of the yeast to the substrate is high, and vice versa. A theoretical derivation proposed by Wang and Li (2014) suggests that K_c in the Monod model should be multiplied by biomass concentration to account for the effect of cell flocculation and diffusional barriers that arise in high-density cell cultures. In Equation (1a), a first-order cell decay term is added to the Contois model to represent the endogenous decay of biomass.

The rate of substrate uptake is calculated by Equation (1b). Previous studies on *X. dendrorhous* fermentation have concluded that the rate of glucose consumption for biomass maintenance and astaxanthin production is negligible relative to that for cell growth (Liu & Wu, 2008). In contrast, the kinetics of astaxanthin formation have been modeled using the Luedeking–Piret equation (first two terms on the right-hand side) shown in Equation (1c), as previous studies have observed that astaxanthin can be accumulated during both biomass growth (growth-dependent synthesis, first term on the right) phase and biomass stationary phase (growth-independent synthesis, second term on the right) (Luna-Flores et al., 2010). However, the accumulation rate of astaxanthin was also observed to decrease rapidly in the current experiments when biomass concentration is high and glucose is depleted. A decrease in astaxanthin titer at the later stage of fermentation was also observed in some of our previous experimental datasets (Figure S2). This can be explained by the reasoning that cells may reverse astaxanthin into other metabolites for their maintenance as there is no external carbon source available. As a result, an astaxanthin consumption term is added in Equation (1c). This term is assumed to be proportional to X^2 as the decrease of astaxanthin accumulation rate is found more rapidly than that of biomass (which is proportional to X).

Finally, a detailed step-by-step derivation of this kinetic model (hybrid of the Contois and the Luedeking–Piret model) together with its assumptions and practical limitations can be found in Supplementary A for interested readers.

2.2.2 | Mixed sugar kinetic model construction

For mixed sugar cultivations, Equations (1a)–(1c) must be expanded to account for the different rates and efficiencies of the two sugars utilized. Different model structures can be adopted depending on the nature of the multiple-substrate limitation (Bekirogullari et al., 2020). The multiplicative approach assumes that both substrates are essential resources for growth such that the overall growth rate is simultaneously co-limited by the availability of the two nutrients. The additive approach postulates that the substrates are substitutable, hence the availability of each substrate contributes individually to the growth of the organism. The noninteractive approach considers that the substrates are catabolized via independent pathways. Thus, there is no growth co-limitation of the different substrates, rather it is the most limiting nutrient (e.g. the one providing the lowest specific growth rate) that controls the cellular growth.

$$\frac{dX}{dt} = \left(\frac{\mu_{m_1} \cdot S_1}{S_1 + K_{c_1} \cdot X} \cdot \frac{1}{1 + k_1 \cdot S_2}\right) \cdot X + \left(\frac{\mu_{m_2} \cdot S_2}{S_2 + K_{c_2} \cdot X} \cdot \frac{1}{1 + k_2 \times S_1}\right) \cdot X - \mu_d \cdot X, \tag{2a}$$

$$\frac{dS}{dt} = -Y_{S_1} \cdot \left(\frac{\mu_{m_1} \cdot S_1}{S_1 + K_{c_1} \cdot X} \cdot \frac{1}{1 + k_1 \cdot S_2}\right) \cdot X - Y_{S_2} \cdot \left(\frac{\mu_{m_2} \cdot S_2}{S_2 + K_{c_2} \cdot X} \cdot \frac{1}{1 + k_2 \cdot S_1}\right) \cdot X, \tag{2b}$$

$$\frac{dP}{dt} = \alpha_1 \cdot \frac{\mu_{m_1} \cdot S_1}{S_1 + K_{c_1} \cdot X} \cdot \frac{X}{1 + k_1 \cdot S_2} + \alpha_2 \cdot \frac{\mu_{m_2} \cdot S_2}{S_2 + K_{c_2} \cdot X} \cdot \frac{X}{1 + k_2 \cdot S_1} + \beta X - k_d X^2, \tag{2c}$$

where k is a constant accounting for substrate inhibition, and the subscripts 1 and 2 refer to glucose and sucrose, respectively. It should be noted that the sugar concentrations S_1 and S_2 were not measured separately, rather they were lumped in a total sugar concentration S. Other symbols retain the same meaning as in Equations (1a)–(1c).

In this study, glucose and sucrose are considered substitutable substrates as *X. dendrorhous* has been observed to utilize both mono- and di-saccharides carbon sources for growth (Vázquez et al., 1997). This implies that a multiplicative approach would not be appropriate. Furthermore, noninteractive models are only applicable when there is a small degree of interaction between the biochemical pathways of each substrate (Bader, 1978). In the case of X. *dendrorhous*, sucrose is first hydrolyzed into fructose and glucose before being consumed by the cells (Sheu et al., 2013), thus a large degree of interaction between sucrose and glucose is expected. Following this reasoning, we turn our attention to additive models. Additive structures have been previously used to model bacterial growth in the presence of multiple competitive substrates. For instance, Turon et al. (2015) proposed one such model combining the Monod and Haldane kinetics to predict the growth rate of *Chlorella* microalgal species. The additive model in this study expresses glucose-dependent and sucrose-dependent growth of *X. dendrorhous* through two separate Contois functions with substrate inhibition terms.

The inhibition terms (i.e., $\frac{1}{1+k_1 \cdot S_2}$ and $\frac{1}{1+k_2 \cdot S_1}$) are multiplied to the glucose and sucrose Contois equations, respectively, to account for inhibitory effects between the two substrates. These terms have been adapted from the work of (Kwon & Engler, 2005) and (Schmitt et al., 2016). If strong inhibitory mechanisms are present ($k_i \gg 0$) and the inhibiting substrate concentration is high ($S_j \gg 0$), then it can be deduced that $\frac{1}{1+k_i \cdot S_j} \gg S_i$ and the inhibited substrate growth expression tends to 0. Conversely, in the absence of significant inhibitory effects ($k_i \approx 0$) or if the inhibiting substrate is depleted ($S_j \approx 0$), then $\frac{1}{1+k_i \cdot S_j} \approx 1$ and the growth expression for the inhibited substrate is reduced to a classic Contois model.

The parameters associated to glucose-dependent growth (μ_{m_1}, K_{c_1}, Y_{S_1}) were first determined in the single-sugar cultivation case study. This allowed for the estimation of the sucrose-dependent growth parameters, which otherwise would have been nonidentifiable as individual sugar concentrations were not measured in the mixed-substrate cultivations. Furthermore, to evaluate the effect of sucrose on the endogenous decay of biomass and astaxanthin formation and accumulation, μ_d, β, and k_d in the dual-substrate model were also re-estimated based on the mixed sugar experimental data. The reader can refer to Supplementary B for an alternative model formulation that predicts biomass growth, substrate consumption, and product formation to a similar degree of accuracy as the model presented in Equations (2a)–(2c). This alternative model structure incorporates competitive inhibition terms, adapted from the work of Yoon et al. (1977), to account for repression effects between the two substrates. Albeit it was discarded by comparison of the fitting error and number of parameters of the proposed models, we believe this model is also a valid representation of the fermentation kinetics.

2.2.3 | Parameter estimation

A dynamic model parameter estimation problem was formulated as a nonlinear least-squares expression presented in Equations (3a)–(3c). The objective function to be minimized is a least-squares formula (Equation (3a)) subject to the nonlinear process constraints (Equation (3b)) and bounds of state variables and parameters (Equation (3c)).

$$\min \sum_{i=1}^{N} (\boldsymbol{x}_{i,E} - \boldsymbol{x}_{i,M})^T \boldsymbol{\Lambda} (\boldsymbol{x}_{i,E} - \boldsymbol{x}_{i,M}), \tag{3a}$$

$$st.\ \frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}, \boldsymbol{\theta}), \tag{3b}$$

$$\boldsymbol{x}_{lb} \le \boldsymbol{x} \le \boldsymbol{x}_{ub}, \boldsymbol{\theta}_{lb} \le \boldsymbol{\theta} \le \boldsymbol{\theta}_{ub}, \tag{3c}$$

where $x_{i,E}$ and $x_{i,M}$ are experimental measurement and model estimated values of state variables $\boldsymbol{x} = [X, S, P]^T$, $\boldsymbol{x}_{lb}$ and $\boldsymbol{x}_{ub}$ are the lower and upper bound of state variables, respectively, $\boldsymbol{\theta}_{lb}$ and $\boldsymbol{\theta}_{ub}$ are the lower and upper bound of parameters ($\boldsymbol{\theta}$), respectively, $\boldsymbol{\Lambda}$ is the weighting matrix, and N is the number of total data points.

To implement dynamic model parameter estimation, the kinetic models are firstly fully discretised and transformed into a nonlinear programming problem (NLP). To guarantee the high accuracy of model discretization, a fourth order orthogonal collocation over finite elements in time is used as a discretization scheme (del Rio-Chanona et al., 2015). The optimal parameters of the kinetic model are determined by solving the discretized NLP using the state-of-the-art interior-point nonlinear optimization solver IPOPT (Wächter & Biegler, 2006). The execution of parameter estimation in this study is programmed in the Python optimization environment Pyomo (Hart et al., 2012). The computer specification is AMD Ryzen 5 processor, 3.6 GHz, and 16 GB of RAM. The total computation time was 17.3 s, of which 15.8 s correspond to data importation and only 0.8 s correspond to the evaluation of the nonlinear optimization problem.

2.3 | Hybrid model construction

In reality, values of kinetic model parameters also change with respect to time as each of them represents a number of intrinsic metabolic reactions that are lumped into a single kinetic term. As the activity of these metabolic reactions changes due to the dynamic culture environment, the lumped kinetic parameters also have different values over the experimental time course. Using a single set of parameter values results in large uncertainty and low accuracy of a kinetic model, particularly if the structure is non-identifiable. As a result, in this study, a hybrid model that integrates a data-driven model, namely Gaussian processes (GP), within a simple kinetic model is proposed and formulated as Equations (4a)–(4c) for the single sugar experiment. To avoid repetition, we used the single sugar experiment as an example to illustrate the performance of a hybrid model. The same procedure can also be applied to the mixed sugar experiment.

$$\frac{dX}{dt} = \left(\mu_m \cdot \frac{S}{S + K_c \cdot X} - \mu_d\right) \cdot X = \mu\left(X, S\right) \cdot X, \tag{4a}$$

$$\frac{dS}{dt} = -Y_{S/X} \cdot \left(\mu_m \cdot \frac{S}{S + K_c \cdot X}\right) \cdot X = -Y_{S/X}\left(X, S\right) \cdot X, \tag{4b}$$

$$\frac{dP}{dt} = \left(\alpha \cdot \mu_m \cdot \frac{S}{S + K_c \cdot X} + \beta - k_d \cdot X\right) \cdot X = Y_{\frac{P}{X}}(X, S) \cdot X, \tag{4c}$$

By re-arranging the kinetic model, a hybrid model is developed wherein the three data-driven terms $\mu(X, S)$, $Y_{S/X}(X, S)$and, $Y_{P/X}(X, S)$ account for the specific growth, consumption, and accumulation rate of biomass, substrate, and product, respectively. These three terms are constructed using three independent GP models.

2.3.1 | Introduction to Gaussian processes

Gaussian processes (GPs) are a probabilistic machine learning technique, enabling the approximation of an underlying function from a set of data. By first specifying an underlying distribution of functions as prior knowledge and then conditioning this prior distribution with data, the resultant posterior predictive distribution can be inferred. Because GPs enable Bayes rule to be applied to functions, resultant predictions are not single scalar values, but themselves Gaussian distributions, enabling efficient uncertainty approximation. The prior distribution over functions is specified using a mean function, commonly set to 0, and a covariance function, often the squared-exponential function denoted as Equation (5).

$$k(x, x^*) = \sigma^2 \exp\left(-\frac{(x - x^*)^2}{2l^2}\right) + \delta_{ij}\sigma^2_{\text{noise}}, \tag{5}$$

where x and x^* are two input locations, $k(x, x^*)$ is the covariance between them, σ, l and σ_{noise} are hyper-parameters, and δ_{ij} is the Kronecker delta function (Rasmussen & Williams, 2006). Therefore, we specify a prior GP prediction as Equation (6a), and the posterior distribution, following the application of Bayes rule to the prior is as Equations (6b)–(6d).

$$f(x) \sim GP(0, k(x, x^*)), \tag{6a}$$

$$f(x) \sim N(\mu, \Sigma), \tag{6b}$$

$$\mu = K(\boldsymbol{X}^*, \boldsymbol{X})K(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}, \tag{6c}$$

$$\Sigma = K(\boldsymbol{X}^*, \boldsymbol{X}^*) - K(\boldsymbol{X}^*, \boldsymbol{X})K(\boldsymbol{X}, \boldsymbol{X})^{-1}K(\boldsymbol{X}, \boldsymbol{X}^*), \tag{6d}$$

where $\boldsymbol{X}$ is the set of training data inputs, $\boldsymbol{X}^*$ is the set of inputs to be evaluated, K is the gram-matrix of the covariance function evaluated using either the training data and the test data, or solely the training or the test data, and $\boldsymbol{y}$ is the set of associated outputs of the training data (Rasmussen & Williams, 2006). For more information into the implementation and theory of GPs we would guide the reader towards (Bradford et al., 2019; Rasmussen, 2004).

2.3.2 | Gaussian process model construction

In this study, the three GP models have the same input variables (i.e., concentrations of biomass and substrate as derived in Equations (4a)–(4c)). They are used to estimate the specific rates within each time interval. To build the three GP models, initially, kinetic model parameter estimation was implemented to calculate specific biomass growth rate $\mu_i(X, S)$, specific substrate consumption rate $Y_{S/X,i}(X, S)$, and specific product accumulation rate $Y_{P/X,i}(X, S)$ at each time step i for each data set. Once completed, they were used to construct the three GP models (i.e., $\mu(\boldsymbol{x}) \sim GP_1(\boldsymbol{x})$, $Y_{\frac{S}{X}}(\boldsymbol{x}) \sim GP_2(\boldsymbol{x})$, $Y_{P/X}(\boldsymbol{x}) \sim GP_3(\boldsymbol{x})$, where $\boldsymbol{x} = [X, S]^T$) by using an exponential kernel function. The length scale parameters in these GP models were also tuned and their impact on the accuracy of the hybrid model will be discussed in Section 3. Specifically, as most of the experimental data were measured once per 12 h, the time step used to generate training data (i.e., kinetic model parameter estimation) was set as 12 h. However, as bioprocess kinetics evolves more slowly than a chemical reaction, it is possible to assume that within a short time period there is no significant change of kinetic parameters within a bioprocess. Thus, when using the hybrid model to predict unknown process behavior, it is possible to update the values of kinetic parameters from the GP models more frequently. Validation of this assumption will be examined and discussed in Section 3.

2.4 | Model uncertainty estimation

As bioprocesses are generally less reproducible than a chemical process, a high-fidelity model should not only accurately predict the expected process trajectory, but also have low uncertainty for its parameters (thus accurately capturing the process uncertainty). This is of particular importance if robust optimization is to carry out for long-term bioprocess optimization and control. To estimate the uncertainty of the kinetic model for bioprocess dynamic simulation, initially, confidence intervals of kinetic model parameters must be calculated. The covariance matrix for kinetic parameters was approximated by the inverse of the Hessian matrix at the optimal

solution. Confidence intervals of kinetic parameters were subsequently obtained from the trace of this approximated covariance matrix following a standard procedure (del Rio-Chanona et al., 2015; Franceschini & Macchietto, 2008). Once calculated, 100 Monte Carlo sampling tests were carried out to sample 100 combinations of parameter values from their respective distribution. These combinations were finally used to generate 100 process trajectories to approximate the model's predictive uncertainty.

To estimate the uncertainty of the hybrid model, a similar procedure was conducted. However, as a GP model can directly predict the mean and variance of each parameter, 100 Monte Carlo tests can directly sample 100 combinations of parameter values at each time step and use these combinations to generate possible process trajectories without any approximation. As a result, estimating the uncertainty of a hybrid model is more straightforward than that of a kinetic model. In this study, the numerical integration was carried out using CasADi's "CVODES" integrator, and Numpy was used to obtain pseudo-random samples from the parameter distributions for both kinetic model and hybrid model. The performance of the two models regarding uncertainty propagation will be thoroughly compared in Section 3.

3 | RESULTS AND DISCUSSION

3.1 | Results of kinetic model construction

Tables 1 and 2 present the estimated kinetic parameters for the single- and dual-substrate models (Equations (1a)–(1c) and Equations (2a)–(2c)), respectively, and Figure 1 shows the model fitting results. From the figure, it is observed that the kinetic models can well fit both scenarios. The mean absolute percentage error (MAPE) for biomass, substrate, and astaxanthin is 9.03%, 16.9%, and 8.68% for the single sugar experiments, respectively, and 13.1%, 8.00%, and 8.11% for the mixed sugar experiments, respectively, suggesting a good overall fitting performance.

3.2 | Effects of sucrose on bioprocess kinetics

X. dendrorhous biomass growth was best identified by the Contois expression for both single-sugar and dual-sugar cultivations, suggesting that the addition of sucrose does not significantly affect biomass growth kinetics. A possible explanation is the high initial weight ratio of glucose to sucrose (2:1) in the mixed substrate experiment, meaning that glucose is the preferred carbon source. However, the large value of the glucose inhibition constant (k_1 = 5.8) compared with that of sucrose ($k_2 \approx 0$) from the kinetic model suggests that sucrose is the preferred substrate as its presence will suppress of the consumption of glucose. Indeed, this conclusion has been previously reported by another study (An et al., 2001) where sucrose was observed to be the first choice consumed by *X. dendrorhous* for biomass carbon assimilation.

TABLE 1 Single substrate kinetic parameters estimation result

Kinetic parameters	Value	Units
μ_m	0.43	h^{-1}
K_c	63.7	Dimensionless
μ_d	2.10×10^{-3}	h^{-1}
Y_S	2.58	$g \cdot g^{-1}$
α	0.0	$mg \cdot g^{-1}$
β	0.236	$mg \cdot (g \cdot h)^{-1}$
k_d	6.48×10^{-2}	$mg \cdot l \cdot g^{-2} \cdot h^{-1}$

TABLE 2 Dual substrate kinetic parameters estimation result

Kinetic parameters	Value	Units
μ_{m_1}	0.43	h^{-1}
μ_{m_2}	0.132	h^{-1}
K_{c_1}	63.7	Dimensionless
K_{c_2}	3.68	Dimensionless
k_1	5.8	$l \cdot g^{-1}$
k_2	0.0	$l \cdot g^{-1}$
μ_d	5.5×10^{-3}	h^{-1}
Y_{S_1}	2.58	$g \cdot g^{-1}$
Y_{S_2}	1.71	$g \cdot g^{-1}$
α_1	0.0	$mg \cdot g^{-1}$
α_2	0.0	$mg \cdot g^{-1}$
β	0.21	$mg \cdot (g \cdot h)^{-1}$
k_d	4.66×10^{-2}	$mg \cdot l \cdot g^{-2} \cdot h^{-1}$

Nonetheless, the current study also observed that when sucrose is the only carbon substrate for fermentation (Figure S3), biomass growth terminated at a much earlier stage and the Contois expression does not fit the biomass growth curve (indicating a different growth kinetic performance). This suggests that glucose still dominates the biomass growth kinetics in the mixed sugar experiment, even though it is not the preferred substrate. A further literature review reveals that sucrose is not directly metabolized by the yeast, rather it is first hydrolyzed intracellularly into fructose and glucose which are then utilized for growth (Sheu et al., 2013). Furthermore, *X. dendrorhous* cannot effectively utilize fructose as its assimilation rate is much slower than that of glucose (An et al., 2001). Thus, one hypothesis why glucose controls the biomass growth kinetics in the mixed sugar experiment is that once cells consumed the glucose hydrolyzed from sucrose, they can immediately consume the glucose from the culture to keep a high carbon assimilation rate in the mixed sugar experiment.

This hypothesis is further supported by the fact that diauxic shifts can be observed in the experimental growth curves in Figure 1a,b for the glucose and mixed-sugar cases. Given that *X. dendrorhous* is a

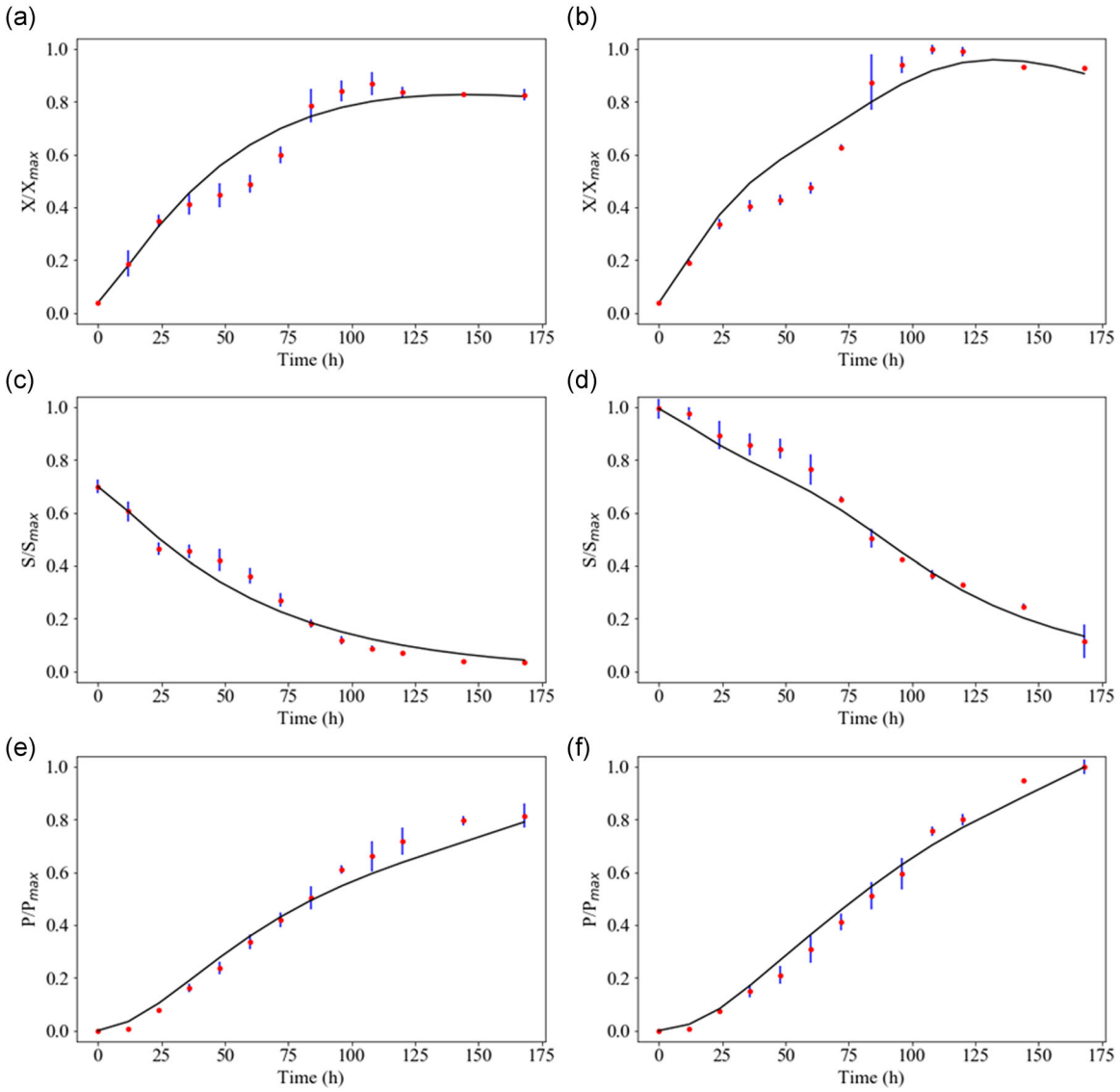

FIGURE 1 Kinetic model data fitting result. (a), (c), and (e) single sugar process; (b), (d), and (f) mixed sugar process. Points are experimental data, blue bars are measurement standard deviation, black lines are model simulation result. The y-axis scales have been normalized against the maximum observed value of the state variables (X_{max}, S_{max}, P_{max}) across both experiments

crabtree positive yeast, the excess of readily metabolizable sugars is associated with overflow metabolism and the accumulation of ethanol in the medium (Reynders et al., 1997). This results in a diphasic mode of consumption where the sugars are assimilated first, followed by the uptake of ethanol upon the depletion of sugars (Lodato et al., 2007). Nonetheless, this behavior was not observed when sucrose was the sole carbon source, as cells cannot assimilate fructose fast enough to activate the overflow mechanism. As a result, when there is glucose available in the culture, cells will prioritize consuming glucose rather than assimilating fructose once sucrose is depleted.

In terms of astaxanthin synthesis, the growth-associated product formation coefficient was found to be 0 $mg \cdot g^{-1}$ for both substrates, indicating that astaxanthin production was growth-independent in both the single-sugar and mixed-sugar cultivations. This agrees with the experimental observations of Lodato et al. (2007) and Wozniak et al. (2011), where carotenoid synthesis by *X. dendrorhous* was induced only once fermentable sugars were depleted towards the end of the exponential growth phase. Furthermore, the growth-independent yield coefficient for the single-substrate case (β = 0.236) is similar to that for the dual-substrate cultivation (β = 0.21), suggesting that the addition of sucrose does not directly influence secondary metabolite synthesis. On the other hand, there is a significant reduction of the astaxanthin-specific consumption rate k_d in the presence of sucrose. This can be justified with reference to the repression mechanism introduced by sucrose, which inhibits the uptake of glucose and may therefore induce adverse environmental stresses like osmotic gradients due to the high sugar concentration in the medium (An et al., 2001), thus promoting the accumulation of secondary metabolites. Enhanced astaxanthin accumulation under unfavorable environmental conditions has been reported for other

carotenoid-producing species such as *H. pluvialis* (Aflalo et al., 2007). However, this benefit is offset by the higher biomass decay specific rate, which causes a more rapid reduction in biomass concentration in the later stage of fermentation. As a result, the optimal concentration of sucrose should be identified in future studies.

3.3 | Comparison between kinetic model and hybrid model

In the current study, significant batch-to-batch variation is observed during the experiments. For example, for the single sugar scenario, as shown in Figure 2, biomass growth and astaxanthin accumulation exhibit different behaviors in the parallel experiments under the same culture environment. The same level of variation is also found in the mixed sugar experiments. As a result, although the kinetic model can well fit the overall trend of the process, its parameters suffer from large standard deviations (e.g., β = 0.236 ± 31.0, K_C = 63.72 ± 27.1) causing the model to have high uncertainties when simulating long-term bioprocess dynamics. This issue has also been highlighted by other researchers in their recent work (Sadino-Riquelme et al., 2020). Figures 3a, 3c, and 3e show the uncertainty propagation result for the kinetic model when simulating the single sugar scenario. This wide uncertainty band prohibits applications of the kinetic model for process robust optimization (e.g., worst-case scenario optimization), as the optimal solution could be over-conservative and has little practical value.

In constrast, from Figures 3b, 3d, and 3f, it is observed that the hybrid model accurately fits biomass growth and glucose consumption in the two experiments. It is worth highlighting that this fitting performance does not indicate overfitting which is often discussed for regression-based machine learning models such as artificial neural network (ANN) (Murphy, 2012). GPs are interpolation-based machine learning models, meaning they are designed to pass through most of the training data points. Overfitting primarily refers to fitting data noise. In this study, all the data used for model construction was averaged over three replicates. Thus, experimental measurement noise has been removed and was not involved in model construction (there is no overfitting). In addition, as seen in Figure 1, the average measurement noise for biomass is 2.46% of biomass concentration, and that for substrate and astaxanthin is 2.99% and 3.16%, respectively. This noise cannot explain the apparent change of state variables during batch cultivation. The highly nonlinear process dynamics exhibited by the averaged experimental data are caused by the complex process kinetics. The fact that the hybrid model can well describe this dynamics indicates that hybrid model is an efficient tool for bioprocess modeling.

Moreover, GPs are more efficient for small data problems and have been used as a better choice over ANNs (Tulsyan et al., 2018). There is only mild model-data mismatch observed at the end of the astaxanthin synthesis profiles. This is because concentrations of biomass and glucose during the later stage of the two experiments are highly similar (thus the same inputs for the GP), but astaxanthin concentrations in the two experiments are different. It is not possible for a model to output two results given the same input. Most importantly, the uncertainty of the hybrid model is lower that of the kinetic model during most of the period of the process. From the figure, it can be seen that most of the uncertainty predicted by the hybrid model is much closer to the real process measurement uncertainty. In addition, the hybrid model can also successfully simulate the diauxic shift biomass growth behavior which cannot be captured by the kinetic model (Figure 3a,b). As mentioned, in reality, kinetic model parameters change over time due to the change of underlying metabolic activity (e.g., a diauxic shift caused by the use of different carbon sources). However, it is not possible to develop a kinetic model to simulate such changes and meanwhile remaining a simple structure. The fact that the hybrid model shows a higher accuracy, lower uncertainty, and better representation of the process dynamics directly speaks of its practical benefits for bioprocess predictive modeling. Table 3 compares the simulation errors between the kinetic model and the hybrid model. More validation results of the hybrid model as well as its advantage over the kinetic model can be found in Supplementary C (Figure S4).

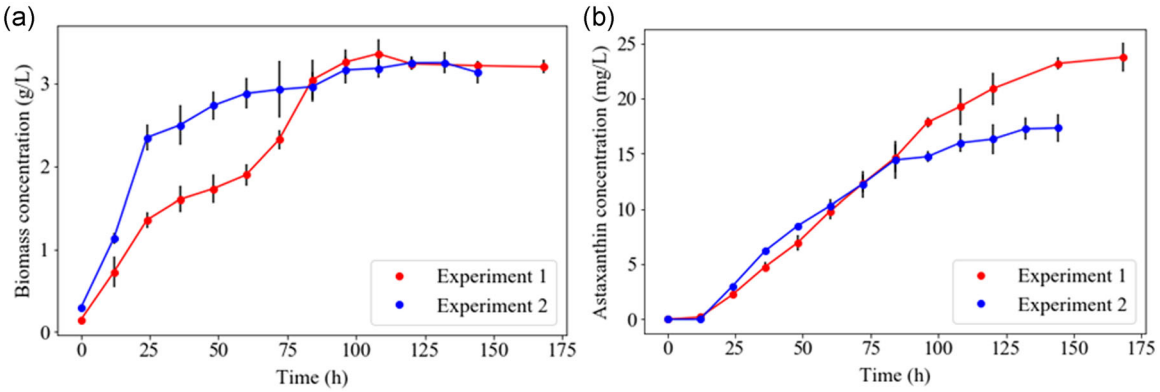

FIGURE 2 Parallel experiments for the single sugar scenario under the same operating conditions. (a): biomass growth in the two runs; (b): astaxanthin accumulation in the two runs. Black bars are measurement standard deviation

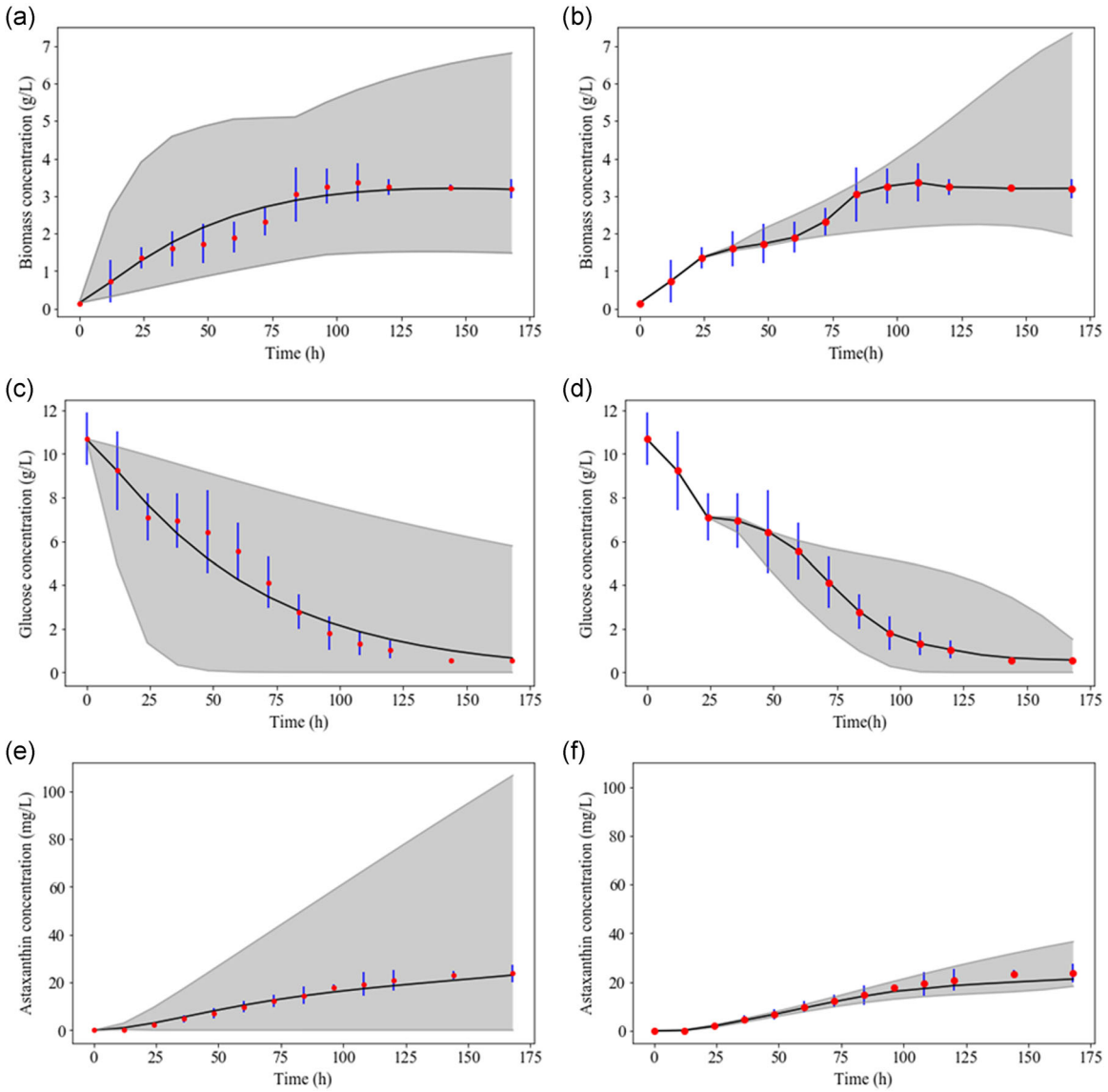

FIGURE 3 Comparison of the kinetic model and the hybrid model. (a), (c), and (e): kinetic model simulation result; (b), (d), and (f): hybrid model simulation result. Points are experimental data, lines are model simulation results, and gray bands are the model uncertainty (99% confidence interval). Blue bars are three times measurement standard deviation (equivalent to 99% confidence interval)

TABLE 3 Mean absolute percentage error (MAPE) and mean percentage standard deviation (MPSD) of different models. *X*, *S*, and *P* are biomass, sugar, and astaxanthin, respectively

	MAPE (*X*)	MPSD (*X*)	MAPE (*S*)	MPSD (*S*)	MAPE (*P*)	MPSD (*P*)
Kinetic	10.6	32.5	22.1	48.8	8.73	50.0
Hybrid	0.053	15.5	2.14	36.5	7.88	13.2

3.4 | Key hyper-parameters in the hybrid model

To analyze the performance of the hybrid model, a thorough investigation was carried out. It was found that there are two hyper-parameters that greatly affect the accuracy and reliability of the hybrid model. The first hyper-parameter is the length-scale parameter in the GP models. This parameter determines how much interdependence is believed to have between adjacent data points. Decreasing the length-scale parameter causes the covariance to be lower between adjacent locations in the input domain (i.e., lower interdependence). By reducing the scale of this hyper-parameter, the hybrid model can perfectly fit the two conflicting astaxanthin accumulation profiles (Figure 4). This is because the hybrid model assumes that process kinetics under two highly similar (but not identical)

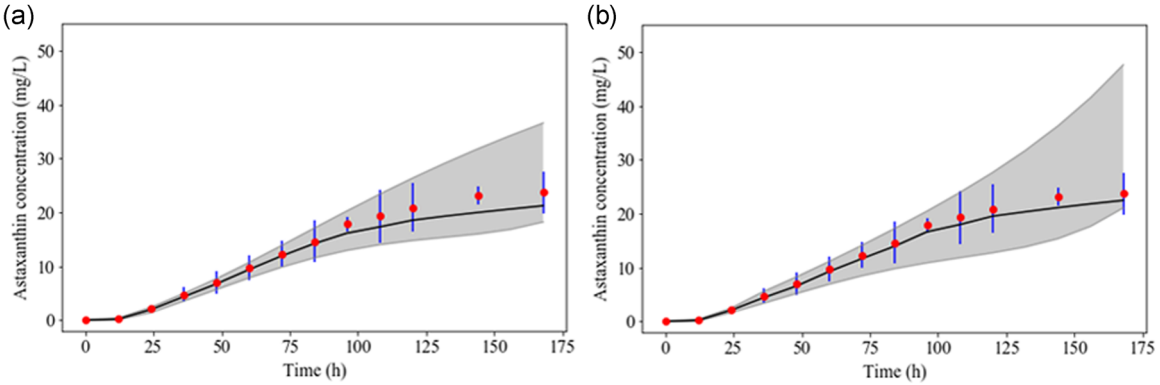

FIGURE 4 Effects of length-scale parameter on the hybrid model's accuracy. (a): length-scale parameter is 1.46; (b): length-scale parameter is 0.20. Points are experimental data, lines are model simulation result, and gray bands are the model uncertainty (99% confidence interval). Blue bars are three times measurement standard deviation (equivalent to 99% confidence interval)

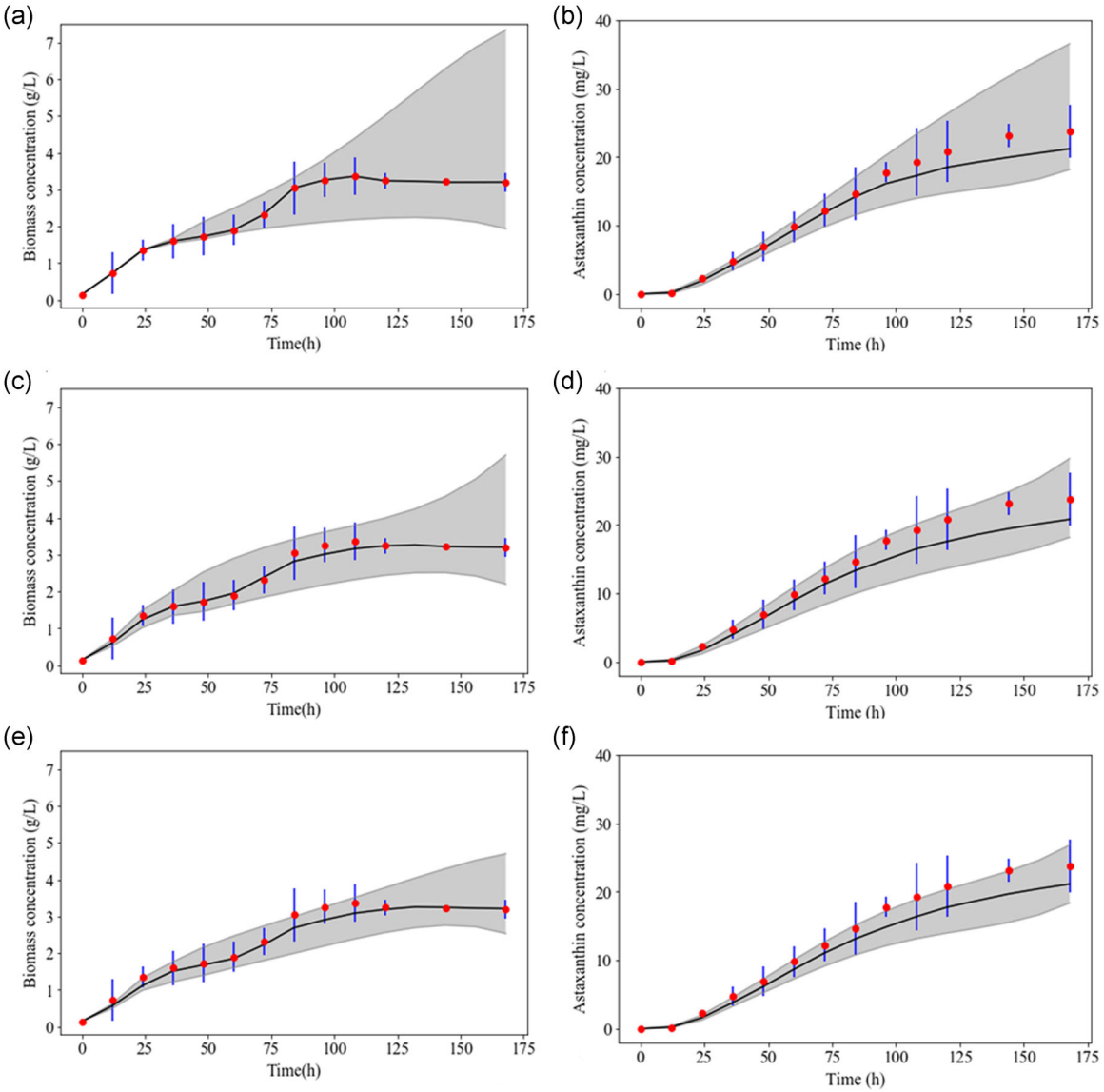

FIGURE 5 Effects of the model update frequency on the hybrid model's accuracy. (a), (c), and (e): hybrid model simulation result on biomass with a model update frequency once per 12 h, 6 h, and 3 h, respectively; (b), (d), and (f): hybrid model simulation result on astaxanthin with a model update frequency once per 12 h, 6 h, and 3 h, respectively. Points: experimental data, lines: model simulation result, gray bands: model uncertainty (99% confidence interval). Blue bars are three times measurement standard deviation (equivalent to 99% confidence interval)

culture conditions are independent from each other and can behave differently. However, the low length-scale parameter amplifies the model uncertainty as the interdependence between adjacent input locations is reduced to minimum. As shown in Figure 4, a smaller length-scale reduces the astaxanthin simulation error from 7.88% to 6.41% (reduced by 18.7%), but increases the model standard deviation from 13.2% to 18.4% (increased by 39.4%). There exists a trade-off between the uncertainty and fitting accuracy of the hybrid model. As a result, the tuning of this hyper-parameter should be cautious.

The other key hyper-parameter is the frequency parameter used to determine how often the hybrid model should update its kinetic parameters (i.e., GPs). Although the current GPs are constructed using average kinetic parameter values within each 12 h, given the assumption that bioprocess kinetics evolves slowly, it is possible to update the kinetic parameters more frequently when using the hybrid model for process simulation. For instance, Figure 5 compares the accuracy and uncertainty of the hybrid model when the GPs are updated once per 3 h, 6 h, and 12 h, respectively. It is found that model simulation errors are increased (from 0.053% to 4.82% for biomass, from 7.88% to 13.6% for astaxanthin) with an increasing update frequency. This is expected as average kinetic parameters will change due to the shortened time interval. The shorter the time interval is, the larger the deviation from the experimental data will be. However, the hybrid model's uncertainty (e.g. percentage standard deviation from 15.5% to 8.60% for biomass, from 13.2% to 9.17% for astaxanthin) is also decreased due to the more frequent update of parameter values. This is because uncertainty propagation is alleviated if the hybrid model can frequently synchronize its parameters. As a result, the frequency parameter also controls the trade-off between hybrid model's accuracy and reliability.

It is critical to emphasize that if the frequency parameter is too high, then the narrow uncertainty of the hybrid model will not be able to reflect the real process uncertainty. For example, when the frequency is chosen as once per 3 h (Figure 5e,f), it is observed that some experimental data has already hit or even cross over the uncertainty bound. If the frequency parameter continues to increase, the model will predict that some of the current process behaviors are unlikely to happen (lying outside the 99% probability distribution), which is also not correct. As a result, a high-frequency parameter underestimates the true process' uncertainty and deteriorates the model simulation accuracy.

4 | CONCLUSION

Overall, based on the current study, it is concluded although the addition of sucrose suppresses glucose uptake, it does not affect the overall biomass growth kinetics. In addition, the use of sucrose can enhance astaxanthin accumulation, but its concentration should be further optimized. Furthermore, it is observed that the Gaussian process embedded hybrid models can capture the dynamics and variability of the underlying process significantly better than the kinetic model. By carefully optimizing the hyper-parameters of the hybrid model, it is possible to reduce the model uncertainty and simulation error by over 60% (well aligned with real process measurement uncertainty) compared with the kinetic model. Using this model for the optimal design of experiments and process real-time monitoring could be more reliable. Future research will focus on identifying the optimal sugar feeding strategy using the proposed hybrid model.

ACKNOWLEDGMENTS

The authors are grateful for the financial support received from the National Natural Science Foundation of China (No. 21776232 and No. 21978244).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Fernando Vega-Ramon https://orcid.org/0000-0001-7684-681X
Thomas R. Savage https://orcid.org/0000-0001-8715-8369
Panagiotis Petsagkourakis https://orcid.org/0000-0002-2024-3371
Keju Jing http://orcid.org/0000-0002-9055-4781
Dongda Zhang http://orcid.org/0000-0001-5956-4618

REFERENCES

Aflalo, C., Meshulam, Y., Zarka, A., & Boussiba, S. (2007). On the relative efficiency of two- vs. one-stage production of astaxanthin by the green alga *Haematococcus pluvialis*. *Biotechnology and Bioengineering*, *98*(1), 300–305.

An, G.-H., Jang, B. -G., & Cho, M.-H. (2001). Cultivation of the carotenoid-hyperproducing mutant 2A2N of the red yeast *Xanthophyllomyces dendrorhous* (Phaffia rhodozyma) with molasses. *Journal of Bioscience and Bioengineering*, *92*(2), 121–125.

Bader, F. G. (1978). Analysis of double-substrate limited growth. *Biotechnology and Bioengineering*, *20*(2), 183–202.

Bekirogullari, M., Figueroa-Torres, G. M., Pittman, J. K., & Theodoropoulos, C. (2020). Models of microalgal cultivation for added-value products: A review. *Biotechnology Advances*, *44*, 107609.

Birol, G., Kirdar, B., & Onsan, Z. I. (2002). A simple structured model for biomass and extracellular enzyme production with recombinant *Saccharomyces cerevisiae* YPB-G. *Journal of Industrial Microbiology and Biotechnology*, *29*(3), 111–116.

Bradford, E., Imsland, L., Zhang, D., & Chanona, E. A. del R. (2020). Stochastic data-driven model predictive control using Gaussian processes. *Computers & Chemical Engineering*, *139*, 106844.

Cabaneros Lopez, P., Udugama, I. A., Thomsen, S. T., Roslander, C., Junicke, H., Iglesias, M. M., & Gernaey, K. V. (2021). Transforming data to information: A parallel hybrid model for real-time state estimation in lignocellulosic ethanol fermentation. *Biotechnology and Bioengineering*, *118*(2), 579–591.

Carinhas, N., Bernal, V., Teixeira, A. P., Carrondo, M. J., Alves, P. M., & Oliveira, R. (2011). Hybrid metabolic flux analysis: Combining stoichiometric and statistical constraints to model the formation of complex recombinant products. *BMC Systems Biology*, *5*(1), 34.

Chi, S., He, Y., Ren, J., Su, Q., Liu, X., Chen, Z., Wang, M., Li, Y., & Li, J. (2015). Overexpression of a bifunctional enzyme, CrtS, enhances astaxanthin synthesis through two pathways in *Phaffia rhodozyma*. *Microbial Cell Factories*, *14*(1), 90.

del Rio-Chanona, E. A., Dechatiwongse, P., Zhang, D., Maitland, G. C., Hellgardt, K., Arellano-Garcia, H., & Vassiliadis, V. S. (2015). Optimal operation strategy for biohydrogen production. *Industrial & Engineering Chemistry Research*, *54*(24), 6334–6343.

Destro, F., Facco, P., García Muñoz, S., Bezzo, F., & Barolo, M. (2020). A hybrid framework for process monitoring: Enhancing data-driven methodologies with state and parameter estimation. *Journal of Process Control*, *92*, 333–351.

Franceschini, G., & Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, *63*(19), 4846–4872.

Geinitz, B., Rehmann, L., Büchs, J., & Regestein, L. (2020). Noninvasive tool for optical online monitoring of individual biomass concentrations in a defined coculture. *Biotechnology and Bioengineering*, *117*(4), 999–1011.

Hallenbeck, P. C. (2012). In P. C. Hallenbeck (Ed.), *Microbial technologies in advanced biofuels production*. Springer.

Hart, W. E., Laird, C., Watson, J.-P., & Woodruff, D. L. (2012). *Pyomo: Optimization modeling in Python* (Vol. 67). Springer.

Kwon, Y. J., & Engler, C. R. (2005). Kinetic models for growth and product formation on multiple substrates. *Biotechnology and Bioprocess Engineering*, *10*(6), 587–592.

Liu, Y.-S., & Wu, J.-Y. (2007). Optimization of cell growth and carotenoid production of *Xanthophyllomyces dendrorhous* through statistical experiment design. *Biochemical Engineering Journal*, *36*(2), 182–189.

Liu, Y.-S., & Wu, J.-Y. (2008). Modeling of *Xanthophyllomyces dendrorhous* growth on glucose and overflow metabolism in batch and fed-batch cultures for astaxanthin production. *Biotechnology and Bioengineering*, *101*(5), 996–1004.

Lodato, P., Alcaíno, J., Barahona, S., Niklitschek, M., Carmona, M., Wozniak, A., Baeza, M., Jiménez, A., & Cifuentes, V. (2007). Expression of the carotenoid biosynthesis genes in *Xanthophyllomyces dendrorhous*. *Biological Research*, *40*(1), 73–84.

Lorenz, R. T., & Cysewski, G. R. (2000). Commercial potential for *Haematococcus microalgae* as a natural source of astaxanthin. *Trends in Biotechnology*, *18*(4), 160–167.

Luna-Flores, C. H., Ramírez-Cordova, J. J., Pelayo-Ortiz, C., Femat, R., & Herrera-López, E. J. (2010). Batch and fed-batch modeling of carotenoids production by *Xanthophyllomyces dendrorhous* using *Yucca fillifera* date juice as substrate. *Biochemical Engineering Journal*, *53*(1), 131–136.

Murphy, K. P. (2012). *Machine learning a probabilistic perspective*. The MIT Press.

Pinto, J., de Azevedo, C. R., Oliveira, R., & von Stosch, M. (2019). A bootstrap-aggregated hybrid semi-parametric modeling framework for bioprocess development. *Bioprocess and Biosystems Engineering*, *42*(11), 1853–1865.

Ramkrishna, D., & Song, H.-S. (2012). Dynamic models of metabolism: Review of the cybernetic approach. *AIChE Journal*, *58*(4), 986–997.

Rasmussen, C. E., & Williams, C. K. I. (2004). *Gaussian processes in machine learning* (pp. 63–71). The MIT Press.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning (adaptive computation and machine learning)*. The MIT Press.

Reynders, M. B., Rawlings, D. E., & Harrison, S. T. L. (1997). Demonstration of the Crabtree effect in *Phaffia rhodozyma* during continuous and fed-batch cultivation. *Biotechnology Letters*, *19*(6), 549–552.

Rodríguez-Sáiz, M., de la Fuente, J. L., & Barredo, J. L. (2010). *Xanthophyllomyces dendrorhous* for the industrial production of astaxanthin. *Applied Microbiology and Biotechnology*, *88*(3), 645–658.

Sadino-Riquelme, M. C., Rivas, J., Jeison, D., Hayes, R. E., & Donoso-Bravo, A. (2020). Making sense of parameter estimation and model simulation in bioprocesses. *Biotechnology and Bioengineering*, *117*(5), 1357–1366.

Schmitt, E., Bura, R., Gustafson, R., & Ehsanipour, M. (2016). Kinetic modeling of *Moorella thermoacetica* growth on single and dual-substrate systems. *Bioprocess and Biosystems Engineering*, *39*(10), 1567–1575.

Sheu, D.-C., Chang, J.-Y., Chen, Y.-J., & Lee, C.-W. (2013). Production of high-purity neofructooligosaccharides by culture of *Xanthophyllomyces dendrorhous*. *Bioresource Technology*, *132*, 432–435.

Tulsyan, A., Garvin, C., & Ündey, C. (2018). Advances in industrial biopharmaceutical batch process monitoring: Machine-learning methods for small data problems. *Biotechnology and Bioengineering*, *115*(8), 1915–1924.

Turon, V., Baroukh, C., Trably, E., Latrille, E., Fouilland, E., & Steyer, J.-P. (2015). Use of fermentative metabolites for heterotrophic microalgae growth: Yields and kinetics. *Bioresource Technology*, *175*, 342–349.

Ukibe, K., Katsuragi, T., Tani, Y., & Takagi, H. (2008). Efficient screening for astaxanthin-overproducing mutants of the yeast *Xanthophyllomyces dendrorhous* by flow cytometry. *FEMS Microbiology Letters*, *286*(2), 241–248.

Vatcheva, I., de Jong, H., Bernard, O., & Mars, N. J. I. (2006). Experiment selection for the discrimination of semi-quantitative models of dynamical systems. *Artificial Intelligence*, *170*(4–5), 472–506.

Vázquez, M., Santos, V., & Parajó, J. C. (1997). Effect of the carbon source on the carotenoid profiles of *Phaffia rhodozyma* strains. *Journal of Industrial Microbiology & Biotechnology*, *19*(4), 263–268.

von Stosch, M., Oliveira, R., Peres, J., & Feyo de Azevedo, S. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, *60*, 86–101.

Wächter, A., & Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, *106*(1), 25–57.

Wan, M., Hou, D., Li, Y., Fan, J., Huang, J., Liang, S., Wang, W., Pan, R., Wang, J., & Li, S. (2014). The effective photoinduction of *Haematococcus pluvialis* for accumulating astaxanthin with attached cultivation. *Bioresource Technology*, *163*, 26–32.

Wang, Z.-W., & Li, Y. (2014). A theoretical derivation of the Contois equation for kinetic modeling of the microbial degradation of insoluble substrates. *Biochemical Engineering Journal*, *82*, 134–138.

Willis, M. J., & von Stosch, M. (2017). Simultaneous parameter identification and discrimination of the nonparametric structure of hybrid semi-parametric models. *Computers & Chemical Engineering*, *104*, 366–376.

Wozniak, A., Lozano, C., Barahona, S., Niklitschek, M., Marcoleta, A., Alcaíno, J., Sepulveda, D., Baeza, M., & Cifuentes, V. (2011). Differential carotenoid production and gene expression in *Xanthophyllomyces dendrorhous* grown in a nonfermentable carbon source. *FEMS Yeast Research*, *11*(3), 252–262.

Xu, P. (2020). Analytical solution for a hybrid Logistic-Monod cell growth model in batch and continuous stirred tank reactor culture. *Biotechnology and Bioengineering*, *117*(3), 873–878.

Yoon, H., Klinzing, G., & Blanch, H. W. (1977). Competition for mixed substrates by microbial populations. *Biotechnology and Bioengineering*, *19*(8), 1193–1210.

Zhang, D., Dechatiwongse, P., del Rio-Chanona, E. A., Maitland, G. C., Hellgardt, K., & Vassiliadis, V. S. (2015). Dynamic modelling of high biomass density cultivation and biohydrogen production in different scales of flat plate photobioreactors. *Biotechnology and Bioengineering*, *112*(12), 2429–2438.

Zhang, D., Savage, T. R., & Cho, B. A. (2020). Combining model structure identification and hybrid modelling for photo-production process predictive simulation and optimisation. *Biotechnology and Bioengineering*, *117*, 3356–3367.

Zhang, D., Wan, M., del Rio-Chanona, E. A., Huang, J., Wang, W., Li, Y., & Vassiliadis, V. S. (2016). Dynamic modelling of *Haematococcus pluvialis* photoinduction for astaxanthin production in both attached and suspended photobioreactors. *Algal Research*, *13*(12), 69–78.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Vega-Ramon, F., Zhu, X., Savage, T. R., Petsagkourakis, P., Jing, K., & Zhang, D. (2021). Kinetic and hybrid modeling for yeast astaxanthin production under uncertainty. *Biotechnology and Bioengineering*, 1–13. https://doi.org/10.1002/bit.27950